

This document is for internal Xerox use only.

"GYPSY" AN INVESTIGATION OF THE GINN COMPUTER-ASSISTED EDITING SYSTEM

by Betty Burr and Ralph Kimball

September 29, 1976

Abstract

In April, 1976, Ralph Kimball and Betty Burr of the PARC User Sciences Group conducted an on-site investigation of the computer-assisted editing system at Ginn & Co. in Lexington, Massachusetts.

This investigation studied the effects of the introduction of a computerized system upon *naive* professionals and the effectiveness of the system in addressing the needs of professional editors. Interviews with the users also yielded a taxonomy of textbook editing tasks and suggestions and implications for further research in computerized editing systems.

Keywords: Gypsy, Editing, Textbook Publishing, Word Processing

© Copyright 1976 by Xerox Corporation

XEROX

PALO ALTO RESEARCH CENTER
3333 Coyote Hill Road / Palo Alto / California 94304

This document is for internal Xerox use only.

OUTLINE

I. Introduction	1
A. Reasons for the investigation	
B. Procedures of the investigation	
II. Rationale for a computer-assisted editing system	3
III. Taxonomy of tasks in textbook publishing	6
A. General list and description of tasks	
B. Ginn editors' involvement in specific tasks	
IV. Evaluation of Gypsy and suggestions for Cypress implementation	21
A. User reactions and suggestions for Cypress implementation	
B. Location of system and system management	
C. Orientation, training and support	
V. Implications for future research in computer-assisted editing systems	32
A. Possible engineering additions	
B. Document composition research-- Inputting, Visualization, and Annotation	
VI. Summary	42
VII. Appendices	44
A. List of Ginn users interviewed	
B. Sample manuscript sheet	
C. Sample art specification sheet	
D. Sample of manuscript specifications	
E. Sample pages from ISIS mini-text	

I. INTRODUCTION

Reasons for the Investigation

Gypsy is the software portion of a PARC-developed editing system currently in place at Ginn & Co., a Xerox-owned elementary/high school textbook publishing house in Lexington, Massachusetts. The system currently consists of three Alto mini-computers, a Diablo Hi-Type printer, a Redactron composer, and the Gypsy software. The complete system is referred to at PARC as *Gypsy* and at Ginn as the *CAE*.

The development and installation of Gypsy was originally proposed by Ginn management as a means of reducing publishing costs. To encourage the development of necessary software, Ginn hired Timothy Mott to work in conjunction with Larry Tesler of PARC. Delivery and installation of the system at Ginn was accomplished in March, 1975.

Since that date, between 15 and 20 Ginn employees have been initiated into use of the system. Two booklet texts, a systems management guide, and a teacher's guide have been completed on the system. Up to 10 additional titles, ranging in size from 48 page guides to a series revision with 3000 pages are planned for 1976 system production.

Gypsy has therefore become the first PARC developed system to function in a real-user environment outside of PARC. This field placement offered an opportunity for a user study investigation to determine the effect of computer-assisted editing on non-computer professionals.

An investigation was conducted in April, 1976 by Ralph Kimball and Betty Burr of the User Sciences Group at PARC. Ralph Kimball is a computer scientist with an interest in editing systems and in the impact of new technologies on professional settings. Betty Burr is a former editor with Xerox Education Publications, publishers of supplementary text materials.

The objectives of the investigation were:

- ** To determine what effect the introduction of computer technology had upon the working patterns of the naive users at Ginn.
- ** To discover whether Gypsy was meeting the needs of Ginn editors and to determine the value of individual system features.

- ** To bring back suggestions from Ginn users concerning changes to or expansions of the editing system for possible implementation in Cypress, the follow-on computer based editor.

Procedures of the Investigation

From April 20th to 22nd, the PARC team conducted intensive interviews with sixteen Ginn employees. Eight of those interviewed were editors with anywhere from 5 and 25 years editorial experience at Ginn who had used Gypsy. One editor with no Gypsy experience was interviewed.*

Three secretaries were questioned. One woman, a new employee, was being trained on the system. The other two secretaries were experienced at Ginn and had worked on the system with one or more manuscripts.

The method of investigation was individual interviews, with one or both investigators questioning the users. On the final day, a group of three editors met with the investigators.

A brief questionnaire concerning specifics of the Gypsy system was read to most interviewees. During the major part of each interview users were encouraged to independently voice their reactions to the system. Editors were asked to discuss their entire editorial process step by step, including their use of Gypsy. Secretaries, who do most of the manuscript input on the system, were questioned about their reactions to specific Gypsy features.

Mary Emerson, trainer and operator of the system, spoke with the team at length, as did Darwin Newton, Executive Managing Editor and key person behind the computer-assisted editing program.

The team also observed the training of two new users--one editor and one secretary.

In addition to the interviews and observation of training, the investigators noted the physical placement of the Gypsy system, the management of time on the system, and the use made of the system as indicated in a log book kept in the CAE room.

**A full list of those interviewed, their subject fields, and years of experience may be found in Appendix A.*

II. RATIONALE FOR A COMPUTER-ASSISTED EDITING SYSTEM

The educational publishing industry has been suffering from the general decline in both the publishing industry and the education market. As school enrollments began to shrink, competition among textbook and materials publishers became increasingly severe. At the same time, costs of printing and distribution increased dramatically.

Publishers were forced to evaluate more carefully the profitability of producing any new text. In one case, Ginn management decided not to bid on a new textbook in Texas because the number of publishers planning to bid indicated that Ginn could not expect a reasonable margin of profit for their investment.

The following table of revenue data in the publishing industry presents current industry conditions and shows the position of Ginn in relation to other publishers.

Table 1

Publishing Industry Revenue Data
United States--1975
(all figures in millions of dollars)

Newspaper and periodical sales	\$ 11,000
All book revenues	\$ 3,810
Textbook revenues	\$ 1,170
Elementary/high school textbooks	\$ 527
Ginn & Company revenues	\$ 41 (1.1% of all books).

Size of Ginn content editorial staff: 55 (out of 284 total employees)

Projected total size of all book editorial staffs, (assuming
same ratio of revenues to staff size as Ginn): 5111

Faced with ever increasing costs and competition, textbook publishers must expect a small margin of profit; at Ginn the average after-tax profit is 10%. Under these circumstances, the publisher must shave costs wherever possible. Ginn has already reduced their in-house graphics staff and contracts out most graphics work. The

editorial staff has been cut back considerably from 1974 levels, and all publishing operations previously carried out on two floors of their office building have been consolidated on one floor.

In proposing the computer-assisted editing system as a cost savings measure, Ginn anticipated a cost reduction of from 15 to 20 percent. The system was expected to cut down secretarial typing, editing time, and costs for setting the manuscript into type.

As of April 1976, thirteen months after installation, Ginn had insufficient experience with the system to know whether their projections were correct. This is not surprising, because the number of bugs associated with a research prototype system can be expected to be large. Accurate measurements of productivity could not be expected during the first year, in any case. In addition, no projection of cost savings has yet been made for graphics capabilities of the proposed Cypress system.

The following breakdown of costs for an average textbook, obtained from Darwin Newton, Executive Managing Editor of Ginn, shows areas in which Ginn anticipates a savings with Gypsy.

Table 2
Cost for an Average Textbook

Book: 608 pages; for grade 5-12; 4-color

Editorial/Graphics:	\$100,000 (amortized in first year of book sales)
---------------------	--

Editorial:	\$65,000
<u>Savings with CAE</u>	<u>\$15,000</u>

(Savings anticipated through use of Gypsy to increase productivity or actually save time on editing/typing--based on applying a fixed percentage to editorial costs--not on a work flow itemization)

Graphics:	\$35,000
-----------	----------

Plates: \$100,000 (amortized over a five year period)

Composition \$20,000

(3 stages of proofs:
galley/pages/finals)

Savings with CAE \$ 3,000

(Savings expected with Gypsy-15% through preparation of compositor-ready manuscript or tapes--based on applying a fixed percentage to composition costs)

Art & cover originals \$30,000
(photos/original art/cover design)

Camera & film \$25,000
(mechanicals/negatives)

Stripping \$22,000
(assembly of negative materials for plate making--separate for each color)

Printing plates \$ 3,000
(\$40 per surface--plate prints 32 pages
separate plate for each color of 4-color page
deep edge plates-\$100)

(Total Savings Anticipated: \$18,000)

Printing Costs: \$1.40 per book

average run for high school text--20,000/year

average run for elementary text--100,000/year

Selling Price: \$7.50 net to schools

Profit Margin Expected: 20% before taxes; 10% after taxes

III. TAXONOMY OF TASKS IN TEXTBOOK EDITING

Words could not adequately describe the delicate intricacies of the publishing process. -- a Ginn editor

Computer-assisted editing can decrease textbook publication costs only if the system effectively shortens some of the book production stages. Traditional methods of preparing books for print involve a number of lengthy, *intricate* steps.

In some of these stages, editors may have little or no involvement. Ginn editors estimate that 60% to 90% of their time is spent on actual editing work, while the remainder of their attention is divided among other publishing tasks. The Gypsy system was designed to interface only with editing tasks and must be evaluated on the basis of performance in assisting with those tasks. Nevertheless, an understanding of the editors' work does require comprehension of the total range of publishing activity. This taxonomy of publishing tasks provides an outline of an average publishing cycle.

Within the publishing industry editorial responsibilities will differ with house policy and with the nature of the material being prepared. Therefore, this general industry compilation includes some responsibilities not assigned to all editors at Ginn. Special mention is made of those areas in which Ginn editors concentrate most of their efforts.

The range of publishing activities in this list has been arbitrarily divided into six categories. The order is sequential, although design and editing are overlapping activities. Each category description contains a commentary on the Ginn editors' involvement in the work of that stage. The quotations by Ginn editors came from both the interviews and a report to Ginn management.

The publishing cycle generally consist of:

- A. Rationale and Planning
- B. Design, Layout and Illustration
- C. Manuscript Editing--primary editing, final editing, and proof reading
- D. Printing
- E. Sales and Marketing
- F. Maintenance and Changes to Existing Texts

A. Rationale and Planning

The generation of an idea for a given book or series occurs in one of two ways: a) an idea is submitted to the publishing house by an outside author or group; or b) the house itself determines the need for a new text and seeks authors. In the first case, either the manuscript alone or an entire package including rationale, planning, field testing of lessons, and even layout and design may be submitted by the outside authors. In the second instance, all steps in the publishing cycle are performed in house.

In both cases, before the decision is made to produce the book or series of books, a thorough investigation of the feasibility and profitability of the book is conducted. This Rationale and Planning stage may last anywhere from three months to over a year. Steps that are taken to investigate the feasibility of producing a certain text include the following:

Identify audience	<i>Determine grade level, reading level, and subject area of the target group.</i>
Determine market need for new text	<i>Identify state adoption committees seeking new texts. Determine whether old texts from the publishing house are pedagogically outdated. Specify new subject fields in which the house has no texts.</i>
Investigate competitors' texts	<i>Determine how many existing texts are on the market and whether better texts could be produced.</i>
Study any existing market research, or request new research	
Estimate costs of project	

If the decision is made to produce a book or series, a plan, or *Program Proposal* is developed, including:

Determination and writing of educational objectives of text

Planning of general outline of text

Determination of specifications of text--reading level, rough design, layout (these specifications are sometimes required by state education departments)

Selection of writers and request for sample materials from writers (usually at least one chapter and outline for entire text)

Meetings with writers

Selection of outside consultant experts to evaluate and validate samples and text

Evaluation of sample materials by consultants and senior editors

Once the evaluations have been completed and the decision is made to proceed with the text, work is assigned to an editor who follows the guidelines established in this initial planning stage.

Ginn Editors' Involvement in Rationale and Planning

...In my experience as a Ginn editor [seven years], I have never participated in a 'program planning stage' as so described.

Since editorial will have the closest contact with the author in subsequent stages, editorial should be involved in the planning and formulation of...working procedures.

At Ginn most of the rationale and planning is done by management or by senior editors. This planning process does not, therefore, directly concern many of the editors using the Gypsy system. However, the initial decisions do greatly influence the work of all editors.

B. Design, Layout, and Illustration

Preliminary design and layout of a text usually occurs simultaneously with early manuscript editing. Determinations must be made that create an integration of text and graphics to serve the pedagogical requirements of the book. Early in the production cycle, graphics designers, sometimes in collaboration with the book editor, will produce layout sheets and/or sample pages. Some of the initial considerations are:

- length of text
- percentage of text to illustrations
- placement of illustrations
- style of illustrations--size, color, values, subject treatment
- text type
- headline type
- line count maximums for pages
- page allocations for chapters/units
- preparation of stylesheet

After the preliminary layout is designed, the art is assigned to an illustrator whose style meets the needs of the particular text. Concurrently with the editing process, this artist will submit three stages of artwork for approval by the designer and by the editor, if house policy calls for editorial approval:

- roughs--brief sketches of several alternative treatments
- comprehensives--more detailed renderings of selected treatments
- final artwork--the finished product to be sent to the printer

During the editorial process, editors are specifically concerned with many of the following graphics considerations:

Use stylesheets (preset for the text)	<i>Stylesheets specify decisions about format made in the preliminary design phase and are given to editors by graphics designers.</i>
Prepare manuscript on pre-set sheets	<i>Manuscript sheets indicate number of characters and lines. Editors type to preset limits for individual text fonts. The resultant line count is therefore fairly accurate. (See Appendix B for a sample manuscript sheet.)</i>

Determine amount of white space.	<i>Editors may indicate "white space" in the manuscript in which they wish no illustration or text in order to effect a rhetorical pause.</i>
Indicate style marks; label manuscript	<i>The style marks on the finished manuscript are meant for the compositor who will set the type for the text. The manuscript is labeled to show the approximate position of illustrations.</i>
Specify/interpret specs for illustrations	<i>Indicate to artists through words or sketches the proportions, locations, and content of the illustration most appropriate for the age group, suitable to the text, and pedagogically valuable. (See Appendix C for sample art specification sheet.)</i>
Check pedagogical features of art	<i>Pedagogical value is roughly 60% of the consideration about any artwork or illustration.</i>
Confer with designer on mechanicals	<i>Mechanicals contain pasted-up typeset text and/or graphics and are camera-ready for the printers.</i>

Ginn Editors' Involvement in Design, Layout, and Illustration

...graphics/editorial personnel [should] work together closely on the creation of page layouts, ideally sitting down side-by-side to complete at least the first one or two chapters/units...especially...areas such as science, social studies, and mathematics where placement of graphic material is often crucial to the understanding of text.

The issue of whether length and page position of text is determined by editors or by graphics designers has always been a controversial one in the publishing industry. Ideally, as the Ginn editors suggested in the above quotation, the effort should be a cooperative one. In reality, such cooperation seldom happens.

At Ginn, the responsibilities for graphics and editorial are well divided. While editors do suggest art style and positioning, final layout and format decisions are specified by graphics designers with management approval.

Graphics draws up and sends to the editor specification sheets for the manuscript. In some houses, manuscript is typed on sheets marked with character and line counts to correspond with a given text font. Manuscript sheets are not used by most Ginn editors, although graphics designers have indicated a desire for some preset format to be included in the next stage of the computerized editing system.

An additional factor determines the amount of involvement of both Ginn editors and Ginn graphics designers in decisions about layout and format. If a book or a series has been developed by outside authors, the plans for layout may also be suggested by these experts. One such case is the ISIS series--*Individualized Science Instructional System*--high school mini-texts. These booklet texts have a high pedagogical interdependence of text and graphics. Each booklet has been divided into activities, and each activity designed in facing pages (spreads). Other types of texts are submitted as manuscript only and layout is developed by Ginn personnel. The following chart specifies the differences in method of development and type of material between an ISIS mini-text and a more traditional high school text.

	ISIS Booklet	Psychology Text
<i>Authorship</i>	Written by Florida State Univ.	Text written by outside author
<i>Validation</i>	Tested and validated by NSF	Validated by normal Ginn consultants
<i>Editing</i>	Major changes made before submission to Ginn; minor changes made by Ginn	Put through normal Ginn initial editing process and final editing steps
<i>Layout</i>	Determined by FSU & NSF	Determined by Ginn
<i>Text</i>	50% text, 50% graphics	80% text 20% graphics
<i>&Graphics</i>		

(Sample pages from an ISIS text are shown in Appendix E.)

C. Manuscript Editing

INTRODUCTION

Variations in the source and nature of material in a text determine both the amount of graphics design and the amount and type of editing required of Ginn personnel. Necessary editing work will be of three different types:

primary (initial) editing	- extensive changes to manuscript
final (copy) editing	- changes to wording, style, fine points
proof editing and reading	- final checks on manuscript and/or galleys

In general, some amount of each type of editing is required on every manuscript. The primary editing work, however, is the most extensive and subjective part of the editing process. Primary editing involves considerable change to the structure, content, and flow of the manuscript, and may demand major rewriting of the document. Such editing constitutes from 60 to 90% of the editing work on most manuscripts. Generally, textbook editors do not make such major revisions until the author has first tried to correct some of the problems.

Although textbook editors are not hired as writers, the rationale for major changes to manuscript was described by a panel of Ginn editors in a report to management:

Too often the author does not successfully fulfill his role in creating a complete and viable manuscript, and when revision or rewriting is called for, too often the author fails to comply successfully. The result: too often the publisher (in particular the editor) must take on the role of revising, rewriting, and in some cases creating entirely new manuscript. This is professionally bad news, but it happens frequently.

While the amount of primary editing will vary with the condition in which the manuscript is received and with the editor involved, certain universals exist in the objectives, techniques, and work habits among professional text book editors.

PRIMARY EDITING--initial editing stages

Objectives

Meet educational objectives of text
set in rationale stage

Create concept flow; assure continuity

Text material must be presented in the most useful order and must not create confusion or require excessive rote memorization.

Limit concept load

The concept load is measured by the number of ideas in 1000 words, and by the number and progression of concepts in a chapter. The load must be appropriate to the audience reading level.

Assure clarity of ideas

Clarity is assured by rewriting and tightening sentence structure and by improving word choices. The editor relies upon experience and ability in making judgments about how to improve the text.

Control readability level

Readability is the measurement of the difficulty of the text against the expected reading grade level of students. It is determined by special formulas such as the Fry Formula which makes use of a count of the number of syllables in a 100 word sample. If the level is too high, editors replace difficult words and break up sentences. Some editors do not use formulas but judge readability intuitively.

Assure appropriate styling

Each publishing house has a preferred style. In addition, editors must assure appropriate mention of women and minorities.

Techniques--traditional methods used in primary editing

Conceptualize entire text, and sections	<i>Manuscript is often received in sections or chapters. The editor must judge the progression of material on the basis of what is already on hand and what is expected.</i>
Research subject area	<i>Amount and intensity of research depends upon the condition of the manuscript and familiarity of the editor with the subject area.</i>
Locate anthology material	<i>Such material may be used as examples in the text or may be the entire contents of the book.</i>
Annotate manuscript; return to author	<i>Marginal notes are used to communicate with authors, since authors are first asked to make substantial changes to manuscript before editors solve remaining problems. Marginal notes also serve as communication with secretaries, and as reminders to the editor of additional changes.</i>
Reorganize structure of manuscript	<i>Reorganizing, moving entities, and integrating examples all require constant scanning, movement, and multi-page visualization of text.</i>
Move entities for better flow Compose substitute text- (word choice, tone, mood, style, etc.)	
Integrate examples, explanatory text, additional material	
Check activities for educational objectives	

Develop teachers' manual

All texts have accompanying materials for teacher use. The manuals are developed concurrently with the text and may contain suggestions for supplementary teaching materials, methods for handling the lessons, and questions for students.

Like any other line of work, the profession of book editing is filled with traditional methods of operation passed on from one generation of editors to another. These work habits greatly influence the requirements of the users when they approach a system like Gypsy.

Work Habits

Work under great deadline pressures

Editors are often working under strict monthly, weekly, or even daily deadlines. Pressure to meet deadlines is great, since graphics work, typesetting and press time are pre-scheduled for each manuscript.

Use reference books/libraries

Editors constantly verify text and add background or material from references.

Hold much in head

Editorial judgments are based partly upon reading done before and during editing and upon the editor's experience.

Take work home

Due to heavy deadline pressure and the necessity to use at-work hours on other responsibilities, most professional editors count on doing a portion of the editing work at home.

Work in chapters/sections

These two work habits require the ability to spread out several pages at once and look at them concurrently--the number of pages mentioned most often was 4-6.

Compare old version and new	<i>To see how changes in one part of the text affect other sections in flow, concept, load, etc., editors refer to old versions of the manuscript while creating new manuscript.</i>
Let manuscript sit & return later	
Rework manuscript several times	<i>Primary editing usually requires several passes at the manuscript before the editor is satisfied. The number of passes depends upon how much time the editor has before the manuscript must be turned over to the designer and the typesetter, the quantity of work needed on the manuscript, and the amount of perfection demanded by the editor. Usually, however, editors will continue improving a manuscript until they must turn it over to be typeset.</i>
Make marginal correction notes	<i>Marginal notes, often handwritten, are vital for annotating, querying authors, and indicating corrections. Space on the page must be available for these additions.</i>
Use manuscript paper	<i>Manuscript paper with character and line counts is an aid to help editors judge how many lines they have used against the total line length assigned for a section of manuscript.</i>
Have secretary retype worked-over manuscript for clean second edit	
Finished ms. sent to author for approval	<i>Authors make early manuscript changes but are discouraged from altering finished copy to avoid delays in the book production.</i>
Inform managers of progress	<i>Editors must keep records of the progress on various manuscripts.</i>

COPY EDITING

The second phase of editing work is copy editing--final editing. The scope of these activities is more limited, more objective, and more specific than those of primary editing. Most copy editing is performed after initial or primary editing is completed and major changes to the manuscript have been accomplished. Experienced editors do make some copy editing changes during initial editing stages, altering minor errors as they appear. However, concentration at the earlier stage is not on catching all finish editing details. In some publishing houses special copy editors or copy readers make final checks of the document, but at Ginn the editors themselves are responsible for all such functions.

Check consistency of style

Issues of style include introductions, headings, page references, exercises, examples, and indentation. Publishing house standards demand that ideas of equal importance or value be presented in equal form. Editors will also assure that explication through examples is provided wherever necessary and that the pedagogical value of exercises is appropriate to the reader level and interest.

Check specifics of wording

These specifics include spelling, hyphenation, grammar, preferred house style, typos and choice of words. Editors will often make several passes through a document looking for a few of these specifics each time. While the number of items being searched for differs with each editor, all use scanning and replacing procedures.

Check for flaws in arguments

See if headings are a good summary

A summary reading of chapter or section headings should give the reader an accurate picture of the contents of the text.

Expose gaps/excesses in reasoning

Check Teachers' Guide page references

Page references to teachers' guides cannot be inserted into the guide until the text is typeset and pages for the text are set.

Create bibliography

Create index

Indexing must be done after a text is completed and page positions are determined.

Check typeset specs on manuscript

It is often the responsibility of editors to indicate instructions to the typesetter on the manuscript.

Set tabs or paragraph indents

Position math/science formulas or spelling lists

Indicate special symbols

Indication of page placement for poetry or special text may be the responsibility either of the editor or of the designer, depending upon the procedure of each publishing house.

PROOF READING

During or after the primary and copy editing process, the manuscript is marked with typesetting symbols to indicate proper font information and is then sent to the compositor, or typesetter. The third and final editing function is carried out when the compositor sends back page proofs, *galley*s. This pass is merely a check for accuracy, although some small changes may be made on typos or incorrect line counts.

The cost of making editorial changes increases substantially as the manuscript passes through the three editorial stages. For that reason, editors are heavily discouraged from making changes at the galley stage. Authors who insist upon making changes when they approve galley proofs are often charged with AA's--authors's alterations--and may actually be charged money for the changes.

Final galley check

A check is made for typos, gross errors in placement, font specifications, correct pagination, and correct art positioning. The last four items are usually checked by the graphics people, sometimes with the assistance of the editor.

Confer with designer/illustrator

The inclusion of the editor to decide on the acceptability of the proofs at this stage depends again upon the policy of the publishing house.

Allow author check of proofs, if required

Ginn Editors' Involvement in Manuscript Editing

Ginn editors report that editing tasks make up 60 to 90% of their work load. Of that time, 60 to 90% is spent on primary editing, with the remaining 10 to 30% of the time devoted to copy editing or proof reading. With the exception of using manuscript sheets and performing the indexing function, Ginn editors perform all the tasks listed in this editing section.

The use of manuscript paper is optional at some publishing houses and required at others. Ginn does not require it, and most editors do not make use of these sheets. Houses that do make use of this aid maintain that editors are assisted in determining how much space they have left in a given page allocation by having their manuscript typed to close line counts. Ginn graphics designers have expressed a desire to have preset style sheets on the Cypress system which will serve as an automatic manuscript sheet.

At Ginn indexing is currently done by outside index editors. The job takes 6-8 weeks and a delay can slow down the production of the text.

F. Printing

Using the layout plans, designers paste up corrected compositors galleys and finished artwork on mechanicals, positioning text and graphics to fit the pages. These mechanicals are then *camera-ready*. If printing is to be done by photo-offset process, photographic film is shot from the mechanicals, to be then etched onto printers plates for the book production.

G. Sales/Marketing Involvement

Receive constant input from marketing staff on necessary changes to manuscript

Meet with teachers/school officials

The responsibility for a given text does not end with the preparation of manuscript. Editors are called upon to promote the book, to deal with teachers and sometimes with school officials, and to travel to conventions at which they talk with teachers.

H. Maintenance/ Changes to Existing Text

Change/update text

Text is usually updated each time a new printing of the book takes place. Such update may necessitate changing material that has been deemed too controversial, ineffective, or outdated. This activity is carried on concurrently with other new text responsibilities of the editors.

Prepare copyright updates

IV. EVALUATION OF GYPSY/ SUGGESTIONS FOR CYPRESS IMPLEMENTATION

The Ginn interviews explored the effectiveness of Gypsy in assisting with publishing tasks. Discussions with and observations of the professional editors proved to be a unique opportunity to gather criticisms and insights into how a computer-assisted text editing system should be designed. The findings, of course, were anecdotal and subjective, for there is no established science of observation that leads directly to systems design principles. The evaluations in this section are a compilation of the users' comments and conclusions drawn by the investigators from what was said and implied in the interviews.

When the Gypsy system was first introduced at Ginn, the staff had both high expectations and some anxiety about the new technology. They expected the system to increase their time and work efficiency, since it was in the interest of efficiency that the computer-assisted editing was introduced. Many of the staff members were uninformed about the world of computers and were therefore anxious about problems they might have using the system.

After thirteen months experience with the system, the staff expressed a mixture of enthusiasm and disappointment about Gypsy. Most of those interviewed had used the system only briefly; many had experienced some difficulties with it; however, the majority of users were anxious to continue working on Gypsy.

Universally, the users expressed a desire for the system to be made more reliable. They indicated that the realities of deadline pressures and editing requirements created problems that might not have been experienced in PARC experiments with the system. Users also mentioned difficulties with or confusion about several system features.

In the discussions of the editing process, editors reported that the Gypsy system was very helpful in copy editing work but was not effective for much of the primary editing work. They requested the incorporation of additional features that would respond to editorial needs.

Many of the more detailed observations of the Gypsy user interface were directly relevant to the design of the follow-on system, Cypress.

User Reactions and Suggestions for Cypress Implementation

An evaluation of the effectiveness of the Gypsy system should be restricted to those areas of the publishing process for which it was initially designed--editing and the preparation of compositor-ready text. However, some uses have been found for the system in layout and design.

LAYOUT AND DESIGN

Since the Gypsy system does not have a formatting capacity, it is of limited use in layout and design. At present, the Gypsy editing system assists editors to position white space and indicate space for illustrations. However, since editors habitually draw labeled boxes to indicate art position on the finished manuscript and Gypsy does not have a graphics capability, the boxes must be inserted in hardcopy after editing is finished.

PRIMARY EDITING WORK

Manuscripts requiring heavy primary editing seem unsuitable for Gypsy. Traditional editing techniques and editorial work habits require capabilities not now present in the computerized editing system. Ginn editors had made extensive changes on paper before the manuscript was placed on the system for lesser editing tasks. Manuscripts which required minimal primary editing, like the ISIS mini-texts, have been more fully developed on the system. It should be remembered that Ginn editors estimated that the average text requires 60 to 90% primary editing.

Ginn editors commented on some of the ways in which Gypsy does not allow them to use their traditional working patterns in primary editing.

Extent of Primary Editing Changes and Shortage of Secretarial Help

Many Ginn editors expressed the view that inputting raw manuscript to Gypsy was impractical because of the extensive rewriting expected in the first editing passes, and because of a shortage of secretarial help. Text could be placed on the system for smaller editing corrections once major changes had been indicated on the original manuscript pages by either the editor or the author.

Fear of Losing Work, and Deadline Pressures

We're all under an incredible amount of pressure, but to have to cope with an experimental system, too. . . .

I think I'd throw the bloody machine right out of the damn window if I lost a day's work on it.

Reliability of the system was a very emotional issue with the Ginn editors and was mentioned by most of those interviewed. The editors repeatedly requested a more foolproof way to protect themselves from losing work to system failures in the future. If the editor loses a file, he or she must reconstruct much of the work from memory, since much of the editing process involves making subjective judgments. Loss of work might mean that the editor would be unable to meet a deadline. Since press time, artist time, and release of the book have all been scheduled on the basis of editorial deadlines, such failure is very serious.

Early experiences with the system when some files were actually lost seem to have left editors with an ever-present concern. The degree of concern was inversely related to the amount of experience with the system; increased reliability has convinced continual users that they can have more faith in the system.

Non-portability of System

I like to prop up on a couch and mess around with the manuscript. . . I'd rather be home with a tall drink than be at the office with a machine.

Editors habitually take work home. It is obvious that present computer based editing systems cannot meet this requirement for portability. Only if the system can eliminate sufficient editorial work to preclude the need for home work can it satisfy this primary editing requirement.

Page and Space Limitation Problems

The current system does not satisfy editorial needs to visualize many pages at once, to page through a document very quickly, and to spread out the manuscript and glance through it. While the window feature allows for viewing more than one part of the document at a time, the amount of visible text is inadequate. In addition, most of the editors interviewed did not know how to use the window feature, or had not made much use of it.

The system file folder space limitation also proved to be an annoyance. Users complained about lack of space in the file folders. More than one user had lost work by continuing to input manuscript when files were getting full.

Inability to Make Marginal Notations on the System

Most editors are in the habit of using marginal areas for notations to authors or reminders to themselves or secretaries. At present marginal annotation cannot be done directly on the system. The editor must make rough paper notes while working on the system and later transfer those notes to hardcopy. One editor tried to use italic text to insert comments on the system, but later had to go back and remove all the notes.

Correspondence Between Screen and Final Layout/ Final copy

Users requested a closer correspondence between the page as displayed on the screen and the appropriate page in the final layout. In the initial period of Gypsy use, the system did not even provide for correspondence between hardcopy and screen. This problem has been corrected in a later version of Gypsy.

COPY EDITING

Gypsy takes so much less physical effort to edit.

Working on copy editing tasks involves reading through the manuscript, locating items to be changed, and changing or indicating need for change. With the Gypsy editing system this work is shortened considerably, since all functions of scanning and changing are performed more rapidly on the system than on paper. It seems obvious that the greatest impact of the system is in the work of copy editing. Since the system was initially designed with only these tasks in mind, it can certainly be labelled a success in this area.

However, in addition to the fact that Gypsy does not meet primary editing needs, it does not address several of the copy editing tasks listed by Ginn editors, such as setting formulas, inserting symbols, or specifying columns. Ginn abandoned the idea of creating a spelling text on the system because of the lack of columns. One editor tried to position poetry using Gypsy, but was forced to stop because the hardcopy

version did not show the poetry in the proper place. Several editors commented on the need to do hanging indentations on manuscript, not now possible with Gypsy. It was also universally agreed by users that the ability to produce an index on the system would be particularly valuable.

The editing system also has not been fully used for those tasks for which it is efficient. For example, editors indicated that they seldom used the scan or the substitute features, although these capabilities could be of great help to them in some editing functions.

User comments indicated that the infrequent use of certain features was caused by a lack of understanding of and comfort about these features:

Windows

Most users reported that they did not know how to create windows and did not understand the concept of being able to work on more than one file at a time.

Wastebasket

In Gypsy, the space where all deletions are stored is called the "wastebasket." The name was taken too literally; users did not think of that space as storage of edited material, but rather as garbage that could not be recalled. Few editors made effective use of the wastebasket for storing material to be subsequently pasted back into the manuscript. This was due partly to the awkward nature of wastebasket interactions and the poor visibility of items stored there. The initial size of the wastebasket is only two lines; thus when a large amount of text is cut, only the last two lines of text show. The two line limitation is particularly bad if the editor must scroll the wastebasket or do any manipulations in it prior to pasting. A command to expand the wastebasket was available, but it got lost somewhere between the Gypsy implementers and the editors, as none of them knew about it.

A number of suggestions were made for improving the interaction with the wastebasket. One was to start a new line with each cut so that a succession of short cuts would not be confusingly concatenated on the same line. Another suggestion was to facilitate manipulations involving several consecutive cuts by having a command to automatically group these cuts together. An equivalent idea is to implement an "appending cut" that adds the current cut to the wastebasket selection. Finally, users suggested that a better name be given to the wastebasket, something like "source," or "storage."

Scan

Users indicated little familiarity with this feature. When it was explained to them in the interviews, they decided it would be of limited use but could help them make universal changes in copy editing functions.

Substitute

The nature of this feature was not understood.

Selection

Several users expressed problems with mouse selection. By and large, the older the user, the more trouble he or she had manipulating the mouse. This was especially true when users tried to extend a selection: During the observed training sessions, one older editor who was practicing extended selections failed repeatedly because her hand trembled slightly. The high degree of hand-eye coordination required to extend a selection was exacerbated by the fact that selections cannot be *re*-extended. Thus if the editor's hand moved slightly just before letting up on the mouse button, the selection might be incorrect by a letter, necessitating starting the task over. Even editors who did not experience coordination problems disliked the inability to re-extend a selection.

Text Mode Visualization

The issue of visualization, which came up with the wastebasket, was an important theme that pervaded the impressions drawn from the Ginn editors. For instance, the Gypsy implementers went to great pains to achieve a "modeless" user interface (where the effect of the next keystroke can always be inferred from the document context). Nevertheless, the editors occasionally were surprised by actions taken by Gypsy, apparently because their expectations of the system did not correspond to reality. This was observed most often during type-in with formatting functions such as italicize, make bold, and center. For example, if the user invoked the italic command, and then paused before typing the next letter, there was no indication on the display to remind the user that the next character was about to be treated specially.

File Names

Several editors mentioned that their file system would make more sense to them if

they could edit the file names at will. This appeared to be a system bug rather than an unavailable feature, since the names actually could be edited, but with unpredictable effects on the integrity of the file system. Editors had therefore been instructed that they could not edit their file names.

Formatting

The most important additions to the system suggested by the editors had to do with formatting. Tab and indentation capabilities were universally mentioned. Related to these were suggestions for right flushing, centering with respect to an object, and aligning. Finally, a number of diverse suggestions were made for new editing capabilities that included new character alphabets (e.g. Greek, Cyrillic, mathematics), optional paragraph filling, and a capitalizing command.

COMPOSITOR-READY TEXT

The Gypsy system as delivered to Ginn includes a Redactron composer which prepares tape containing compositor marks. Theoretically, any manuscript placed on the system could be taken off on both Diablo hardcopy and Redactron tape. This tape could then be sent directly to the compositor who would read the tape on his own machines and set the type from this tape rather than from manuscript.

As of April, Ginn's experience with this part of the system had been limited because of technical problems with the Redactron unit.

System Location and Placement of Components

I don't know how other people do it, but I keep track on my calendar of what I do each day, and it's very unusual when I can put down one thing. Usually it's six different things.

When I was being diligent [using Gypsy] my other work fell behind two weeks.

Editors usually interrupt editing tasks to attend to other publishing demands--phone calls, conferences, and other short-term activities. While they are editing, they rely heavily upon resource materials to check against the manuscript.

Since the current Gypsy system is housed in a separate air conditioned room (necessitated by the Altos overheating when they were initially placed out on the operating floor), editors are away from their desks and hence from resource material and from those seeking their assistance. In addition, editors are assigned specific time slots for Gypsy use. Many of the interviewees indicated that they would make greater use of the system if they could turn to it whenever they had free time, just as they now do with a typewriter.

While most of the editors interviewed objected to being away from their desks in order to use the system, the secretaries enjoyed being in a separate room. They appreciated the lack of interruptions by phones and persons.

In addition to the problems associated with having the system in a separate room, users mentioned difficulties with lack of working space, insufficient room to manipulate the mouse at the Alto stations, unnaturally high mounting of the screen, and screen glare.

Orientation, Training and Support

ORIENTATION AND TRAINING

During the interviews with Ginn editors and secretaries the investigators inquired about the training and support provided for users. The team also observed the training of two new users. Since the Ginn system is the first field placement experience for PARC, the feelings of Ginn personnel about orientation, training, and support are very important for future design of appropriate training and orientation.

When the system was originally installed, Ginn users were told that they were getting an experimental system. Interview comments indicated that the users did not fully understand the implications of having a prototype system until they began having difficulties with it. Several editors questioned the fairness of their being asked to use an experimental system in an environment controlled by deadlines and production requirements.

Could Ginn users have been better prepared for the problems they faced? While it is not possible to predict precisely in what ways an experimental system may cause problems, some step of orientation or training should provide enough information

about what to expect so that users do not experience a negative reaction when the system does have problems.

A comment by one of the editors indicates the nature of the confusion and insecurity about the system which still exists.

I feel mystified by the directions Once when I tried to tell the machine something, it asked me a question. I tried to tell it something in English. Then Tim came in and put some jibberish in. The machine responded happily.

What seems to be needed to make users comfortable is some non-technical explanation of how the system functions, a sufficient model so that users generally understand what is happening. Such orientation would produce a greater sense of security and would encourage users to try more system features.

The approach taken to the training so far has been to rely upon a model of the machine as video typewriter and electronic filing cabinet. Such a model serves well to identify functions in terms already familiar to the users and helps them to rely upon their past experience. However, this reliance upon familiar models fails to provide users with an accurate idea of the true functioning of the system or of the advantages of the computerized system over old work methods.

In the current training process, trainer Mary Emerson shows new users the basic operating commands for the system, the method of using the mouse selection device, and techniques for creating and retrieving files. The editorial functions are stressed for new editors; type-in controls are emphasized for new secretary users. In the observed trainings, brief mention was made of some of the lesser used features, but little time was spent demonstrating the use of windows, scanning, or substitute commands. The theory behind the training is to give the users just what they need to use the system immediately without overloading them with options. New users are then encouraged to make use of the system, to learn by doing.

Users apparently did not feel comfortable enough with the system to experiment with some of the features. Mary Emerson's experience also indicated that most of the editor-users did not seek out further explanations of the advanced features either from her or from other users after the initial training sessions. This lack of training may explain the fact that most users rarely used the scan, substitute, or window features. In the interviews, however, users said that they wanted to know more about system features and expressed a desire for an on-line summary of options or a

summary card placed at the Alto station. Some failure of the training or support procedures is implied by this situation.

In addition, when the system failed to respond reliably like video, typewriter, or file cabinet, the users became confused and frustrated. In the process of the training, new Gypsy users are warned to file frequently and to stop working immediately and call for Mary Emerson if the system gives any sign of problems. This admonition is certainly necessary. However, without knowing what is actually happening at such times, users may become uncomfortable about their interaction with the system and feel reluctant to trust their work to it.

Some suggestions for future training on the Cypress system emerged from the Ginn interviews. In addition to a more comprehensive orientation to the computerized system and to a simple model of system functioning, users could benefit from some type of peer training and peer support situation. Since several of the Ginn editors now have sufficient experience to feel comfortable working on the system, these people could be used as a core group to give assistance to newer users. If possible, each new user could be assigned a more experienced mentor. This arrangement would provide the new user with easy access to information about how to perform specific editing functions. It also seems reasonable to tell the experienced users more about how the system functions so that their apprehensions are allayed and they can pass on a sense of confidence to the new users.

Other training materials that might be provided with Cypress and with any new system installation include on-line aids which give detailed information about operations in a specific area. For example, when the users wanted to try various look commands, they could request help which would bring them a detailed summary of only what to do to achieve each of the various looks. Along with an on-line aid, a comprehensive manual with a brief, easy-to-use summary should be available to each user. This summary should be indexed to coordinate the editing requirements of the workers with the capabilities of the system.

Another useful aid would be a video tape showing a user in action, to which users can refer when they feel the need of a refresher or of further explanation. A printed manual of practice sections in real editing problems and a corresponding practice disk would also provide an opportunity for the users to independently improve their skills on the system.

To be effective, these aids must be instantly available to the users at the Alto stations and must be short and easy to use. Time pressures may have been partly responsible

for the fact that most Ginn editors did not return to the trainer for further explanations about the system.

SUPPORT

Another issue mentioned in the interviews was user support. Editors commented unfavorably about sharing Alto disks. At present, the manuscript for each book worked on the system is put on one disk. In the case of a long text or a text series, several editors may be assigned to work on the same text, and therefore will use the same disk. Secretaries will also input manuscript on the same disk assigned to an editor. The editors felt that having other users working with a disk increased the risk of creating disk problems. This feeling added to their sense of insecurity about the reliability of the system.

Both the users and the trainer commented upon the difficulty of obtaining necessary assistance from PARC when system failures occurred. The main problem mentioned was the time difference involved in East-West communications, the fact that during Ginn morning working hours the PARC operation had not yet begun. Delays of several hours in getting assistance were very frustrating to Ginn personnel.

The experience of Ginn users in interacting with PARC raises the issue of the future connection between these two units concerning the Cypress system. The results of this investigation suggest that PARC should be aware of problems in training and support that occur during the use of the system, both as a service to Ginn users and as an important source of information for PARC. One possible method of achieving that interface would be to have a Ginn staff member trained at PARC. This person would become sufficiently familiar with the environment, uses, and problems of the system to be able to deal with any problems that might arise at Ginn. He or she would then be a very effective trainer for Ginn and would be able to keep PARC informed about successes and problems with the system. Development of a curriculum for the training would also produce a prototype training package for future system placements.

V. IMPLICATIONS FOR FUTURE RESEARCH IN COMPUTER-ASSISTED EDITING

Possible Engineering Additions

A set of ideas that emerged from the Ginn interviews poses challenging and unsolved research questions. Comments by Ginn editors were suggestive of document creation capabilities that are not currently being developed by Xerox. Implementation would involve primarily engineering design rather than basic research. These included hyphenation, automatic indexing, and optical character recognition (OCR).

OPTICAL CHARACTER RECOGNITION

During primary editing, the author's manuscript arrives in typed form, is read and annotated, and then often shipped back to the author for revision. There is no easy foothold for the current computer editor in this tight turn-around loop. One of the necessary improvements is having raw manuscript available on the computer. The raw manuscript can now be typed into the system in its entirety, but most editors were not willing to expend that much secretarial effort. A possible solution is automatic machine scanning of the manuscripts--optical character recognition (OCR). OCR technology has acceptable reliability (better than 99% on a word basis) and is not too expensive. However, it has one major drawback for the textbook publishing industry; it requires that the original manuscripts be typed in a special OCR typeface. Since the independent authors comprise a "cottage industry," each author owning his own typewriter, it is hard to imagine how the manuscripts could be produced in OCR format unless the publishers provided the typewriters themselves.

ON-LINE DICTIONARY AND THESAURUS

A dictionary facility would be useful near the end of the manuscript editing phases. Simple misspellings could be corrected by looking up all the words in the manuscript automatically. The dictionary could also contain hyphenation information. The direct additional cost of such a system would be reasonable since a dictionary of moderate size (50,000 words) could run on an unmodified Alto, but there would be some limitations on its use. Without a syntax checking routine, a misspelled word could be passed up if it accidentally spelled some other word (e.g., "there" for "their," "heed" for "need," etc.). These mistakes might be more likely to slip by if editors developed a false sense of security about the program. Also, regardless of the dictionary size, the program will not know some words because they are too specialized

or because they are proper names. Since traditional human proof reading appears to be irreplaceable, the best use of a spelling checker might be *after* the proof reader's check.

A potentially useful facility related to the dictionary service is an on-line thesaurus. During the stages of editing that call for rewording certain sections to fit size constraints or stylistic considerations, a thesaurus would be useful. It is difficult to say how cost effective the thesaurus would be, since most editors have *Roget's Thesaurus* close at hand anyway. However, when editors were required to work at Alto stations away from their desks, a system dictionary and thesaurus service could be attractive, especially if it came at no substantial extra cost.

INDEXING

Ginn usually contracts out preparation of the index for each book. This step in manuscript preparation adds six to eight weeks to the overall production time, since the index cannot be done until after the text pages are set. A computer-assisted indexing facility should be relatively easy to add to Cypress and could save considerable publishing time. The editor could reasonably annotate all the points in the manuscript where an index reference is to go, together with text of the index entry. This could be done on an incremental basis throughout preparation of the manuscript. The actual index would be generated automatically by having the system pass over the entire manuscript collecting all the index entries, noting their page numbers, and sorting them in proper alphabetic order.

SYSTEM PORTABILITY

The final engineering addition of considerable impact would be creation of a portable editing system so the editors could do useful work not only at their desks, but in the reference room of the local library and at home. This notion of portability is part of the design problem currently being investigated by the Learning Research Group at PARC.

Document Composition Research

INPUT OF MANUSCRIPT

A number of interesting new research ideas have emerged since the Ginn interviews were conducted, some as a result of PARC discussions about the Ginn investigation.

These new ideas directly address some of the central concerns of primary editing mentioned by Ginn editors: inputting raw manuscript to the system, effectively visualizing the manuscript in various ways, and having a flexible scheme for annotating the manuscript.

One solution has been proposed for the raw manuscript input problem that eliminates the need for OCR entirely. This proposal could be called an "EFE Cut and Paste" editor, where EFE refers to electronic front end image capture. Devices exist to scan the raw manuscript into the editor's Alto and store a undecoded digital image of each page. This image is an exact duplication of all the *marks* on the paper, as opposed to the usual encoded forms of computer stored text.

The Ginn editor could then scroll through the manuscript on-line, cutting, pasting, and making notes to the author directly on the manuscript. Cutting and pasting would simply amount to shuffling the EFE representation of the manuscript in straightforward ways. No paragraph filling or reformatting would be allowed, but it is not possible to perform these functions with scissors and glue either. Notes to the author could be conventional computer encoded text, typed in by the editor to fit in the left margin of the manuscript. This text would be stored in a separate file, together with a reference to the physical location in the manuscript. When the editor was ready to return the manuscript to the author, the edited image of the manuscript would be transferred back to paper, together with the marginal notes. One major impediment to this idea would seem to be whether the image quality on the screen would be high enough to be read comfortably, and whether the "messing around" with the manuscript, so typical of primary editing, could somehow be done by scrolling through the manuscript on the small Alto screen.

VISUALIZATION

An alternative approach, involving capturing all the keystrokes of the manuscript (either by OCR or manual type-in), directly addresses the challenge of primary editing. This approach is based on both effectively *visualizing* the manuscript and having a flexible scheme for *annotating* the manuscript.

In the world of word processing there has been only one way of viewing a long unstructured document like a manuscript. This has been the conventional full page of text. A few experimental computer based systems (for instance, the RCG editor at the Stanford Research Institute and the Woodstock editor at PARC) have implemented a "summary" mode that collapses the document to show only the first line of each

paragraph. In cases where this first line is a special heading for the paragraph, a rather nice outline of the document suddenly appears on the screen. This outline can then be manipulated and rearranged, with corresponding changes taking place in the complete document. However, since manuscripts do not have this structured format, such a feature would be of limited utility for primary editing.

Primary editing really demands that the manuscript be viewable along various special dimensions, some of which are dependent on the manuscript's content. For instance, in a hypothetical high school textbook on the history of the American West, the overt organization of the book might be by region or by time. The editor might also wish to think of the text as the simultaneous development of a number of interrelated themes such as agriculture, industrial development, the railroads, and social issues like racism. In order to ensure a balanced development, the editor might wish to review the entire manuscript in terms of one or more of these themes.

Even in more mundane concerns about manuscripts such as the measurement of the vocabulary level of the text or the number of new concepts introduced in a given span of words, the editor is striving for a degree of balance in the manuscript that is best revealed by viewing it as a whole. A simple system for viewing the manuscript along various thematic dimensions could produce an overview as in Figure 1. Each vertical strip is a galley-like representation of the manuscript. The outlines of paragraphs in each strip are indicative of actual paragraphs. After choosing a number of themes to view and assigning each of them identifying letters, the editor could manually scroll through the entire manuscript (in normal text display) associating one or more theme identifiers with each paragraph. When the editor invoked the overview command, Figure 1 would appear, complete with the locations of the various themes.

Although the process described in Figure 1 would be useful for understanding the distribution of themes in the manuscript, its real use would be to serve as a set of "pointers" for descending into the detailed text. When the editor selected a certain paragraph symbol, the system would expand the view to encompass the actual text of the paragraph. In this way, the editor could rapidly move through the manuscript, concentrating on the development of a single theme.

Another kind of overview of the manuscript is presented in Figure 2. Here the emphasis is on the separate pages of the final layout. This view emphasizes the relationships of the text to the graphics, and of consecutive and facing pages.

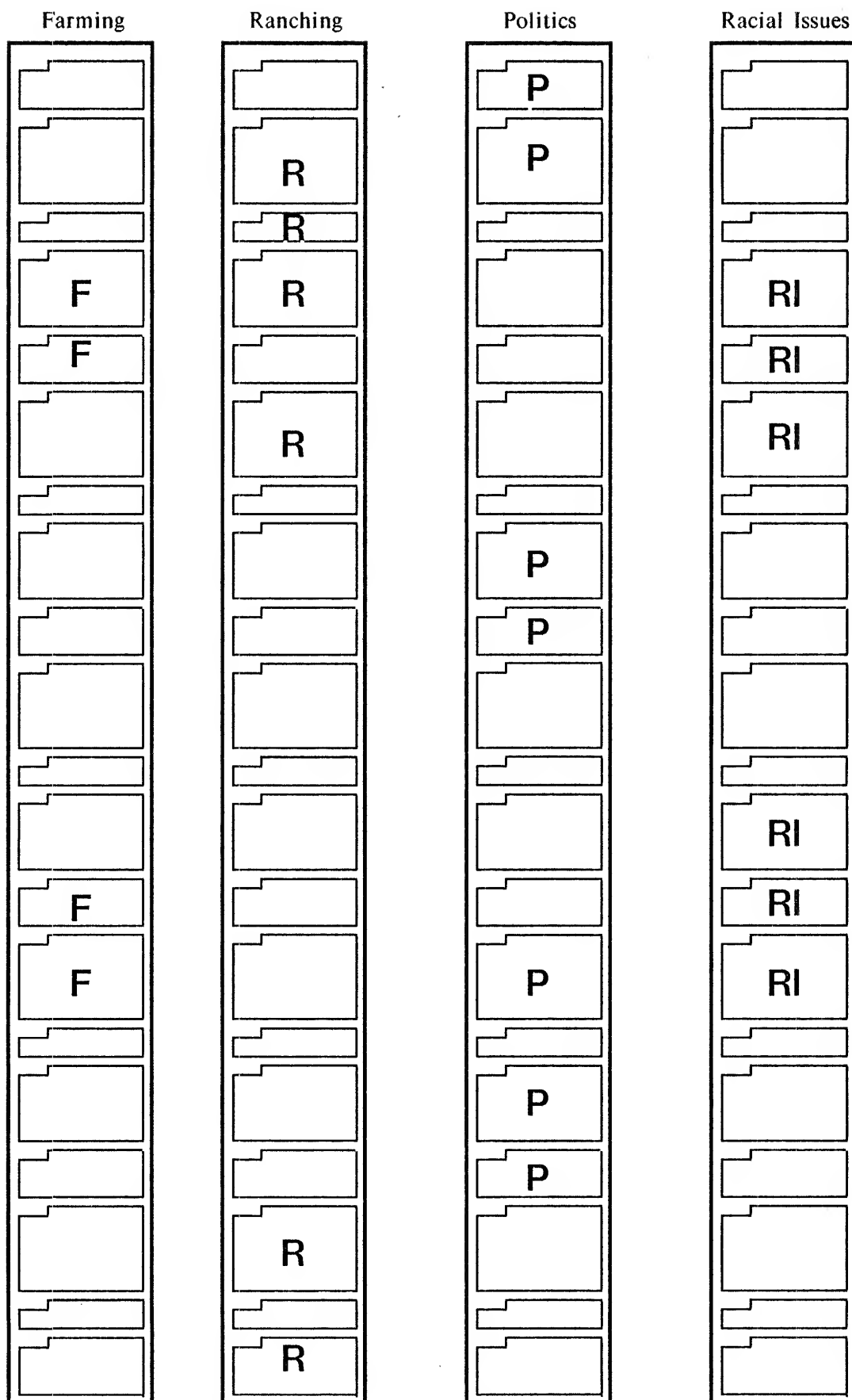


Figure 1. Thematic representation of a manuscript

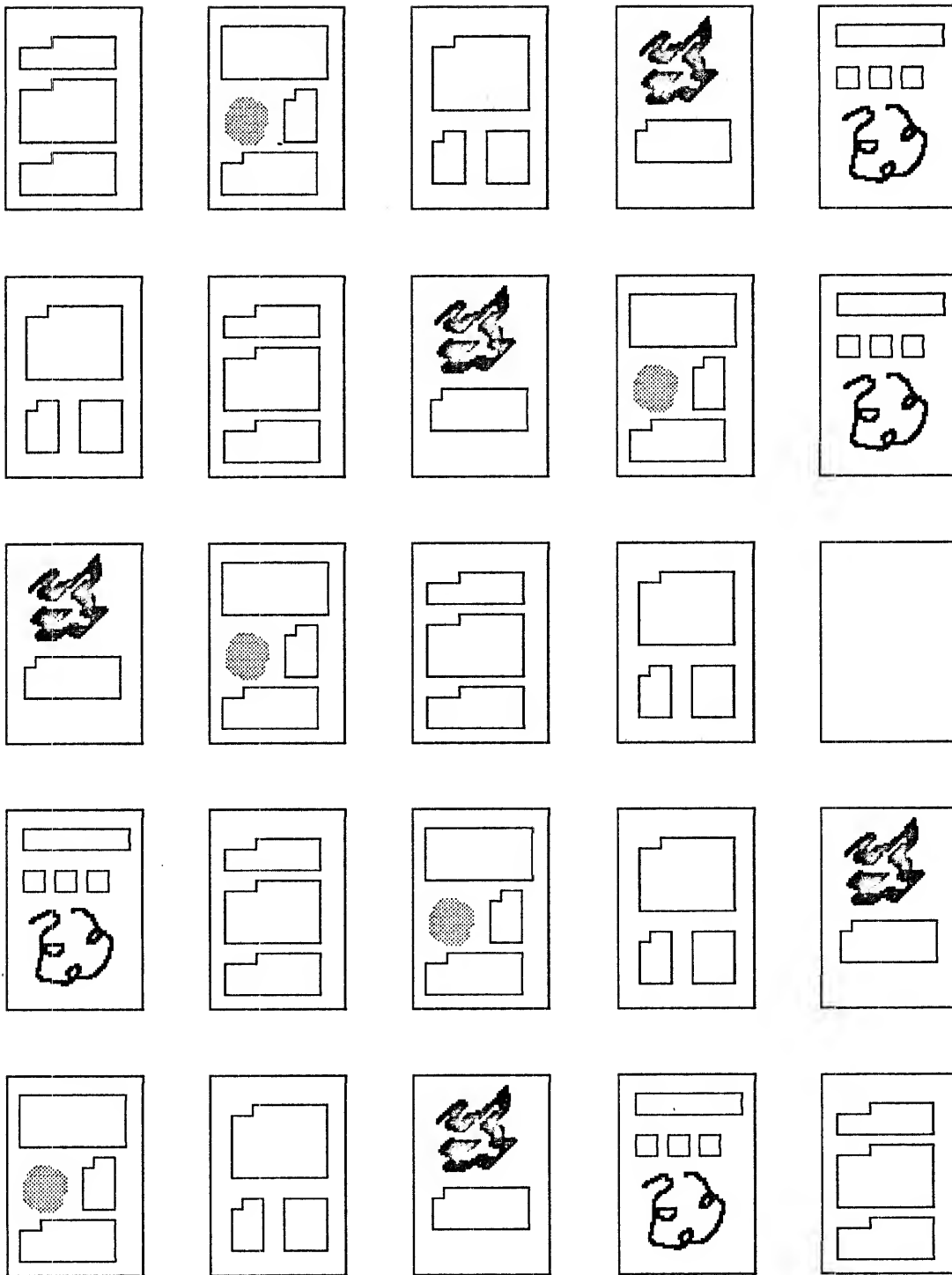


Figure 2. Page oriented representation of a manuscript

An even more exciting possibility involves generating the thematic identifiers automatically. Within the computer science research community, techniques have been developed for lexical analysis of English prose. A fairly straightforward way to generate themes would be to alphabetize and gather together all the separate words of the manuscript. After the removal of a standard set of several hundred filler words (and, of, the, etc.), only words with high thematic significance would be left. Now the editor could scan the word list and pick up groups of words that are related. Each of these groups would define a theme. Since the words came from the text itself, they could then be used as search keys to label the paragraphs with thematic identifiers. Themes that might otherwise not suggest themselves could potentially become apparent. An unresolved question, however, is how difficult or awkward it would be for the editor to combine the words into appropriate thematic groups.

ANNOTATION

Although the mental processes involved in primary editing are probably too complex and subjective ever to be subsumed by the computer, it is possible to augment the computer based editing system in ways that make it an effective *medium* for the human editor. Since it has been repeatedly observed that editors rely on annotations to the manuscript to provide structure for their editing decisions, it seems that an important step in designing a primary editing medium is to provide an annotation capability.

It is crucial to understand that annotations are more than just insertions in the text; they are comments *about the text*. It may also be useful to think of many possible types of annotations, not all of which are of interest at a given moment. For instance, in the real world of Ginn, a manuscript could be annotated in several different ways. One type of annotation might refer to suggested formatting changes, another to stylistic problems, and still another to reminders to think about some readability issue at a later time.

A computer based primary editing system could support a wide variety of annotations. Figures 3 and 4 show simple examples of imagined annotations, where the manuscript is simply drawn on, and text inserted in the left margin or at the bottom. This kind of annotation would be appropriate for the EFE cut and paste editor.

Rewrite:

Here is a line of typical text that is to make up a document. It is typed in an extra small font.
Here is a line of typical text that is to make up a document. It is typed in an extra small font.
Here is a line of typical text that is to make up a document. It is typed in an extra small font.

[illegible]

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

[illegible]

A little terse:

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

Here is a line of typical text that is to make up a document. It is typed in an extra small font.

Figure 3. Example of an annotated document

Here is the most recent version of the paragraph.....

blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3
blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3
blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3
blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3
blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3 blah-3

Here is the 2nd most recent version of the paragraph.....

blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2
blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2
blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2
blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2
blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2 blah-2

Finally, here is the original version.....

blah-1 blah-1 blah-1 blah-1 blah-1 blah-1

blah-1 blah-1 blah-1 blah-1 blah-1 blah-1

blah-1 blah-1 blah-1 blah-1 blah-1 blah-1

Figure 4. Use of annotations to save previous wording

If the manuscript exists in conventional encoded form then a more general annotating scheme can be imagined which would also support visualization of the thematic structure. Present sophisticated document composition systems allow spans of characters within the text to possess various attributes. These are usually formatting attributes like type face, alphabet, and offset. The same mechanism could be extended to associate an annotation with any contiguous set of characters in the text. Ideally, the annotation attribute would consist of an annotation *typeface*, and a *reference* to a file containing the text of the annotation. Specification of the typeface seems desirable so that many different kinds of annotations could be dealt with separately. The existence of a separate file containing the text of the annotation would also be desirable, both to reduce the size and complexity of the manuscript file, and to allow the annotations themselves to be annotated to an arbitrary depth.

With this facility one can imagine many possibilities. Besides the conventional annotations to the author and to the editor, thematic identifiers could be annotations. The previous wording of a sentence or paragraph could also be stored as an annotation, since it might be useful to restore the old wording if the author objected to the editor's changes. A teacher's manual for a textbook would be developed in parallel with the manuscript by storing the text in annotation form. On-line reference material, and even other manuscripts, could be embedded as "research" annotations during the basic development of the text. Finally, one can imagine various system generated annotations sprinkled through the text such as readability measures, creation dates, time when last modified, and the name of the person who did the modification.

VI. SUMMARY

When the computer-assisted editing system was initially introduced, Ginn staff had both high expectations and some reservations. Expectations included ease of operation, cost reduction, and time savings. Reservations were based upon lack of familiarity with the nature of computerized systems and upon fears that use of the system would interfere with their normal operating procedures. Both expectations and reservations seem to have been realized.

Expectations of cost savings have not yet been verified. Time savings has been realized through increased secretarial efficiency and increased ease of copy editing functions. However, the Gypsy system did not provide necessary features to be useful for much of the primary editing work.

Satisfaction with ease of system operation was only partial. Several features of the system were not understood by the users, or did not respond adequately to user needs. In addition, the system has not proved as reliable as users demanded and problems existed with time allotments and with physical location of the system.

Comments by Ginn users indicated specific system features which had proved useful, features which require improvement, and areas in which the system did not meet current editorial needs. Inferences were drawn from these comments about future research issues in document composition, some of which may be addressed by the follow-on Cypress design.

Presentation of some of the findings at a PARC seminar led to follow-up discussions suggesting other research possibilities in computer-assisted editing.

Three important system design principles emerged from the interviews and subsequent discussions:

1. High perceived reliability of the systems is absolutely fundamental for successful field experiments.
2. Complex tasks like primary editing and text-graphics layout require more design attention to *visualization* if the computer is to be an effective medium.
3. A natural extension to the document composition medium is the ability to make *annotations* in a variety of ways.

Observation of training and discussions with the trainer and the users revealed some areas of difficulty with the training and support procedures and suggested methods for developing future training packages.

The Ginn interview project has been a very fruitful "user study." Insights gained from professionals trying to accomplish serious work on a PARC prototype system have been far more valuable than any laboratory simulations of the editing world. The makeup of the interview team was a good example of how such investigations should be conducted. The computer scientist was able to translate the findings into system design metaphors; the professional editor had a sense of what was important and what was hidden from view in the editing environment. The lesson from this experience is that effective evaluation of systems may be greatly enhanced if at least one of the evaluators has professional experience in the target environment.

VII. APPENDICES

A. List of Ginn Users Interviewed

During the three day period the following people were interviewed either individually or in small groups:

Darwin Newton--Executive Managing Editor (He is primarily responsible for initiating the computerized editing concept at Ginn.)

Mary Emerson--Key Operator of System and Redactron.
Trainer on Gypsy

Editors:

Gypsy Users:

Alice Schule--Social Studies--25 years with Ginn

John Bremer--Social Studies--20 years

Janice Ricci--Math/Science--6 years

Rosalee Walters--Science--5 years

Meredith Rutter--Science

Rita Campanella--Science--8 years

Jack Chase--Math/Science

Sam Erskine--Language Arts--8 years

Lucille Waugh--Language Arts--9 years

Non-Gypsy User:

Tom Jackson--Math/Science

Secretaries:

Gypsy Users

Loretta Lewis--Science

Susan Bedard--Math/Science

In addition the team observed the training of two new users, secretary Susan Fenton, and editor Margaret Liggett.

B. Sample manuscript sheet

MASTER MANUSCRIPT FORM

CODE _____ TITLE _____ MS PAGE _____

	45 ELITE	45 PICA
5 ▶		
10 ▶		
15 ▶		
20 ▶		
25 ▶		

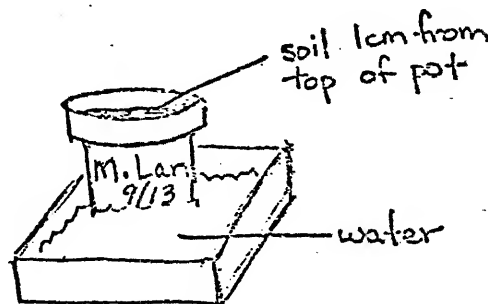
C. Sample art specification sheet

☒ Drawing → ☐ 1/c ☐ 2/c (Color: _____) ☐ 4/c
☐ Photo → ☐ 1/c ☐ 4/c

DESCRIPTION 2

Please illustrate Step A. Show a Mex-Amer male doing the investigation.
Label as shown. *Hand label name & date on pot.*

Art:
The full name
should be written
on the pot -
but it cannot
all be seen.
(The name is
M. Landez)
The date on the
pot is 9/13



D. Sample of manuscript specifications

MANUSCRIPT SPECIFICATIONS

Specimen pages for Purves Project

Attached are specifications for sample pages to be set for the volumes of the Purves Project. An explanation follows:

<u>MS. PAGE</u>	<u>BOOK</u> <u>End</u>	<u>ITEM</u>	<u>CODE</u>	<u>SPECIFICATIONS</u>
1	T12	"FOREWORD"	--	Set for identification
1 and 2	"	Text	A	10 on 11/Galaxy/C & 1c
2	"	Questions	B	12/Galaxy/Caps
3	C12	"Contents"	C	36/Palatino ital./Caps
"	"	"FOREWORD"	D	12/Palatino/Caps
3 and 4	"	Cluster numbers	E	36/Palatino ital./
3	"	Selection titles	D	12/Palatino/Caps
3	"	Selection subtitles	F	12/Palatino/C & 1c
4	"	Selection titles	G	11/Palatino/Caps
4	"	Selection subtitles	H	11/Palatino/C & 1c
3	"	Letter titles	F	12/Palatino/C & 1c
4	"	Letter titles	H	11/Palatino/C & 1c
3 and 4	"	Authors' Names	J	12/Palatino ital./C & 1
3 and 4	"	Page numbers	K	11/Palatino ital./
5	T12	"Contents"	--	Set for identification
"	"	"FOREWORD"	L	11/Galaxy/Caps
5 and 6	"	Part titles	M(1) M(2)	14 Galaxy/Caps 14 Galaxy ital./Caps

DISPLAY TYPE:

1. (Selection Title). From 18 point to 36 point, depending on specific manuscript marking. Set in Caps OR Caps and lower case of the type face designated for the selection that follows. Set solid. Line breaks will be indicated on manuscript. Center in 30-pica-wide type page. If Selection Title is followed directly by start of selection, insert 48 pt. space between.
2. (Selection subtitle). From 12 point to 18 point, depending on specific manuscript marking. Set in Caps OR Caps and lower case of the type face designated for the selection that follows. Set solid. Line breaks will be indicated on manuscript. Center in 30-pica-wide type page. Insert 24 pt. space above, and minimum of 48 pt. space below.

NOTE: During the page layout stage, some Selection Titles and Selection Subtitles will be changed to more "contemporary" or "esoteric" fonts. Ginn will have these lines drawn by hand or set by photolettering methods, and will provide camera copy for shooting and inclusion in 2R page proof.

MAKE UP: Each Core book consists of a differing number of units called "clusters". Each Theme book and Rhetoric book consists of a differing number of unit called "parts". Each "cluster" or "part" to begin a new page, either left or right. "Cluster" word in each Core Book is for identification only, should not be set. Related Arabic numeral is to be set, per specific marking on manuscript. Short sections of typographic matter following the Arabic numeral will be specifically marked for face, size, measure and makeup. "Part" word in each Theme book and Rhetoric book is to be set, with corresponding numeral and Part Title, per specific manuscript marking.

No - are set there

test

field =

12/29/70

BB

which after all is not made of water
asks the water for a face

and anger
different from fire
borrows from it
a loquacious tongue

so is blurred
so is blurred
in me
what white-haired gentlemen
separated once and for all
and said
this is the subject
and this is the object

we fall asleep
with one hand under our head
and with the other in a mound of planets

our feet abandon us
and taste the earth
with their tiny roots
which next morning
we tear out painfully

(30) CC

(40) CC

DD₁

DD₂

EE

Translated by Czeslaw Milosz

and Peter Dale Scott

EE

X

Editor: What did you think of that?

Genteel Lady: Beautiful?

Editor: Beautiful! So, you didn't listen, after
all. I wonder what you'll think of the next
selection. It's about spring in London. But
it's about more than that too.

- Y

Cool Girl - FF

Studs Terkel - GG

Y - This selection is one of the interviews with Chicago people in Mr. Terkel's book, Division Street, America.

Jan Powers, 24 - HH

She's on the staff of a magazine, popular with cool young men. Her job pays well and it's easy. She comes from a working-class family, and occasionally sees her mother and two younger brothers. She has an apartment of her own in a new high-rent high-rise on the Near North Side. She is engaged to Steven, a medical student.

I don't notice the world. I'm very bored. I really don't know how I feel. I'm nice and cordial but people sense something about me. I don't know, maybe I don't like them. Maybe I feel I'm above them. I can't think of anyone I love or respect. I can't be bothered with the news. I just can't get interested. I can't care less. I *should* care, it's terrible. (Laughs lightly.)

Vietnam? Isn't that a shame? (Laughs softly.) I saw a film on Vietnam, it showed the actual fighting. It looked ridiculous, just a bunch of kids. It was actually embarrassing to watch that, people were actually shooting and shouting. I saw Vietnam. I looked at a map once. I'm concerned with Vietnam if my brother has to go, otherwise, no.

My interest in life is me. It's a shame. I wish I could pick up a newspaper and read it. What I hear about things is heard from other people.

I hope I'll make it. I think it's marriage, to someone who is successful. Highland Park, a couple of kids. I'm not too crazy about children, though. You're sitting in a room, and all of a sudden five kids'll come in and they'll go to another girl in the room. Same with dogs.

I'm worried about the next couple of years. Here I'm putting all this time and feeling into this relationship with Steven, and to have it not work out, it would be terrible. I don't know what I'd do. I'd probably find someone else and be just as happy.

Nothing touches me. I wonder why I don't care about these things. The Bomb doesn't bother me. I don't read the papers. There isn't much I can do about it, so I'm not worried. What is important now is my friend and me. The rest of the world can go.

Highland Park: a suburb of Chicago. - LL

STYLESHEET FOR
THE FOLLOWING 18 TITLES OF THE
PURVES LITERATURE SERIES

815100 Core Book 7	815120 Theme Book 7	815140 Rhetoric Book 7
815185 Core Book 8	815205 Theme Book 8	815225 Rhetoric Book 8
815270 Core Book 9	815290 Theme Book 9	815310 Rhetoric Book 9
815375 Core Book 10	815395 Theme Book 10	815415 Rhetoric Book 10
815480 Core Book 11	815500 Theme Book 11	815520 Rhetoric Book 11
815565 Core Book 12	815585 Theme Book 12	815605 Rhetoric Book 12

TRIMMED SIZE: 7-1/2" x 9-1/8". Margins: head, 3/4"; back, 7/8".

MEASURE (MAXIMUM):

- For Grades 7-8: 30 picas wide x 44 picas plus 3 pts. high (exclusive of folio, which is positioned in fore margin, as specified below). Equals 41 lines of basal type.
- For Grades 9-12: 30 picas wide x 43 picas plus 10 pts. high (exclusive of folio, which is positioned in fore margin, as specified below). Equals 44 lines of basal type.

BASAL TYPES:

For Grades 7-8:

- A. 11 pt. Elegante with Italic, leaded 2 points.
- B. 11 pt. Vega with Italic, leaded 2 points.
- C. 11 pt. Zenith with Italic, leaded 2 points.

For Grades 9-12:

- A. 10 pt. Elegante with Italic, leaded 2 points.
- B. 10 pt. Vega with Italic, leaded 2 points.
- C. 10 pt. Zenith with Italic, leaded 2 points.

Prose selections will be identified by single red marginal line on manuscript. Paragraph indentation to be 2 picas. No extra space between paragraphs. Specific measure will be marked on manuscript, as well as whether type is to be justified or ragged right. Maximum single column measure should not exceed 24 picas for grades 7-8, and should not exceed 27 picas for grades 9-12.

Poems will be identified by single blue marginal line on manuscript. Hanging indentation style, with turnover lines indented 2 picas. Specific measure and/or makeup of poem in relation to 30-pica-wide type page will be marked on manuscript. Line numbers (lining figs.), when indicated on manuscript, to be set in 8 point. For poems with short lines, insert 2 pica space between longest line of poem and left side of largest line number; align line numbers vertically. For poems with wide lines and line numbers, maximum width of single line to be 27 picas, then turn over line; line numbers in such poems to be flush right in 30-pica-wide type page. Insert 12 pts. space between stanzas. Insert 12 pts. space between last line and author's name when ms. indicates author's name to follow last line of poem. Make up author's name flush right with longest line of poem.

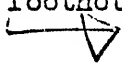
Drama selections will be identified by single green marginal line on manuscript. For grades 7-8, set 27 picas wide, made up flush left in 30-pica-wide maximum measure. For grades 9-12, set 30 picas wide. Hanging indention style, with turnover lines indented 2 picas. Speaker's name or identification to be in cap and small caps, followed by one em space before beginning of statement. Insert 6 pts. space between statements by different speakers. Set stage directions (double green marginal line) block style, 26 picas wide, made up flush right in 30-pica-wide measure. Insert 12 pts. space above first line and below last line of stage directions.

for 7 & 8 (24?)

FOLIOS: 11 pt. Zenith figs., centered north and south on type page, made up in fore margin, center of folio 36½ picas from inside edge of 30-pica-wide type page.

for 7 & 8 too?

FOOTNOTES: Will be identified by double red marginal line on manuscript. Set in 8-point, leaded 1 point, of the font specified for the related selection. Set block style, to the measure specified, made up flush left with the pertinent basal type above. Insert minimum of 12 pts space between last line of selection and first line of footnote. If more than one footnote occurs on a page, insert 4 pts. space between footnotes.



LEGENDS: Will be identified by single brown marginal line on manuscript. Set in 9 pt. Galaxy Light italic, leaded 1 point. Set block style, x 14 or x 28 per specific measure marked on manuscript. Insert 15 pts. space between legend and corresponding photo or art.

STUDENT APPARATUS: Will be identified by single purple marginal line on manuscript. Set in 10 pt. Galaxy Light, leaded 2 points. Set block style OR hanging indention style (2 pica indent) per specific manuscript marking, to maximum measure of 25 picas. Regardless of line lengths, make up left edge of Student Apparatus sections indented 5 picas from left side of 30-pica-wide type page. Spacing between components of a given Student Apparatus section to be indicated on manuscript. Insert minimum of 36 pts. space above first line of Student Apparatus section.

CREDIT LINES: Will be identified by double purple marginal line on manuscript. Set in 7 pt. Galaxy Light, leaded 1 point. Set block style, to maximum measure of 20 picas. Regardless of line lengths, make up Credit Lines at foot of first page of a selection, positioned flush left in 30-pica-wide type page. Insert minimum of 24 pts. space above first line of Credit Lines.

TRANSLATOR'S NAME: Will be identified by a circled cap "T" on manuscript. Set in 8 pt. Galaxy Light italic, made up flush right with right side of author's name above, with 2 pts. space between.

*Not stated above
relative to
manuscript*

E. Sample pages from ISIS mini-text

advanced

Activity 9 Planning

Activity 10 Page 43

Objective 14: Describe the roles of oxygen, glucose, ATP, mitochondria, enzymes, and the Krebs cycle in cellular respiration.

Sample Question: Match each item in List A with its role in cellular respiration in List B.

List A

- oxygen
- glucose
- ATP
- mitochondria
- enzymes
- Krebs cycle

List B

- active acetate enters this
- makes ATP and glucose
- essential to all respiration reactions
- reacts with hydrogen from "hydrogen pathway"
- where most respiration reactions occur
- original source of energy
- energy-storage molecule in cells

Objective 15: Describe the main steps in the cellular respiration of glucose.

Sample Question: What is the order of these steps in the cellular respiration of glucose?

- formation of active acetate
- formation of water
- Krebs cycle
- formation of pyruvic acid

Activity 11 Page 50

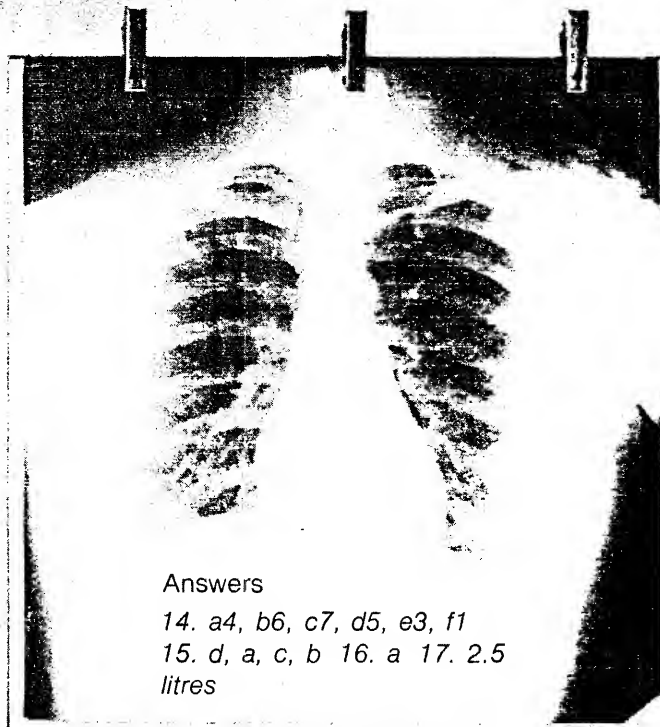
Objective 16: Tell how pressure and volume of a confined gas at constant temperature are related (Boyle's Law).

Sample Question: Assume a constant temperature. If the pressure on a litre sample of air were decreased, the volume would

- increase.
- decrease.
- stay the same.

Objective 17: Solve problems about gas pressure and volume using Boyle's Law.

Sample Question: Suppose the pressure on a 10-litre sample of air were changed from 1 atm to 4 atm at constant temperature. What volume would the air occupy at 4 atm?



Answers

14. a4, b6, c7, d5, e3, f1
15. d, a, c, b 16. a 17. 2.5 litres

★ 7-9. Based on the results in Figure 7-2, which gas seems to control your involuntary breathing rate?

Most people would guess that a low oxygen level is why you can't hold your breath very long. But it's really a buildup of carbon dioxide in your lungs and blood that forces you to breathe again.

You have a breathing control center in your brain (Figure 7-3). It receives information, or *feedback*, from many parts of the body. It senses especially how much carbon dioxide is in your blood.

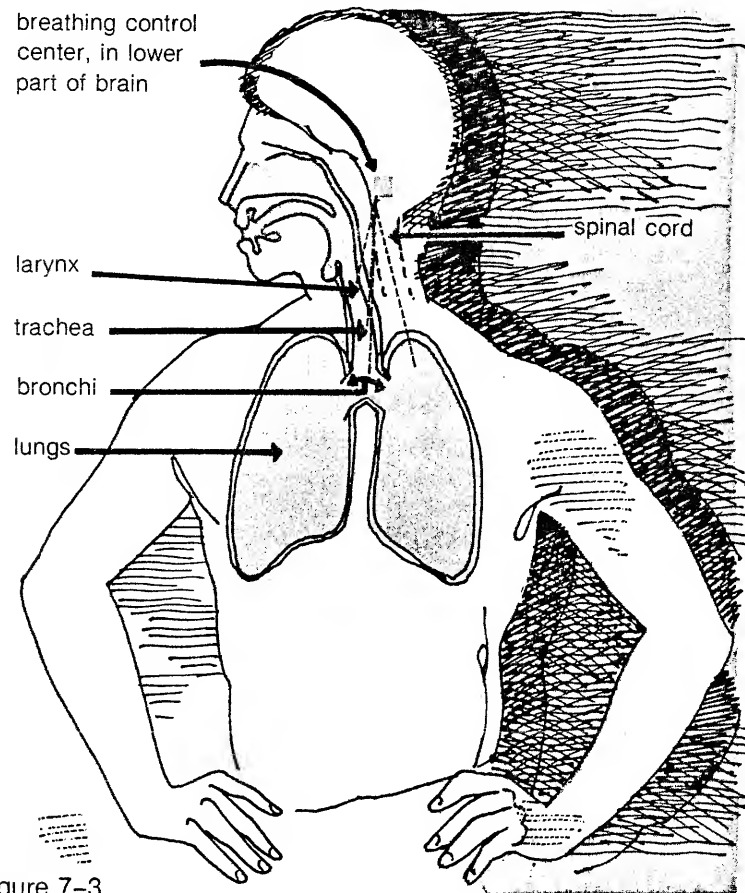


Figure 7-3

Carbon dioxide is in the air you breathe, but it also comes from your body cells as a waste product. When your breathing control center senses more carbon dioxide than normal in the blood, the center speeds up breathing. The increased breathing rate gets rid of the extra carbon dioxide. In this way your breathing system controls itself. Scientists call this kind of self-control *feedback control*.

This may be the first time you've run into feedback control. If so, don't worry if the meaning isn't completely clear. You'll study feedback control in other minicourses. If you've already run into this type of control and still feel uneasy about it, you may want to read *Resource Unit 13*.

Figure 7-4 shows how your breathing control center reacts to feedback. It works to keep the amount of carbon dioxide in your blood constant. Whenever the amount of carbon dioxide (CO_2) changes, the breathing control center responds and adjusts your breathing rate. You can see that the center never rests. It constantly controls the amount of CO_2 in the blood by constantly controlling the breathing rate, whether you're exercising or sleeping.

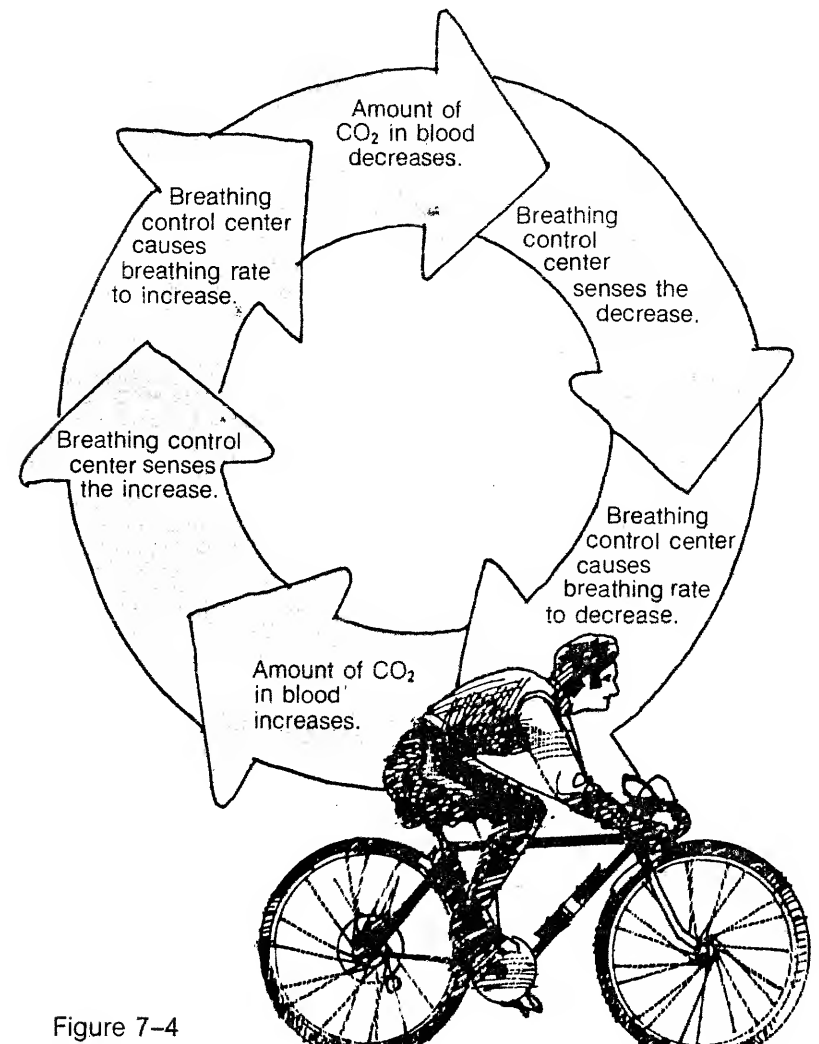
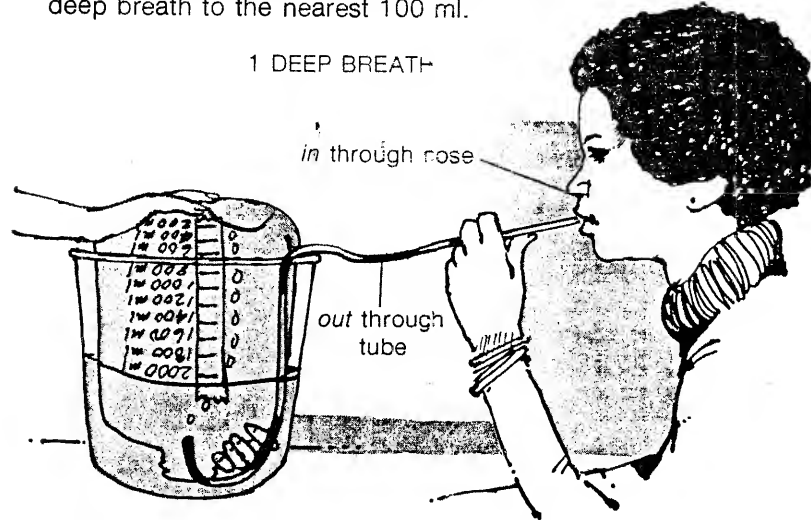


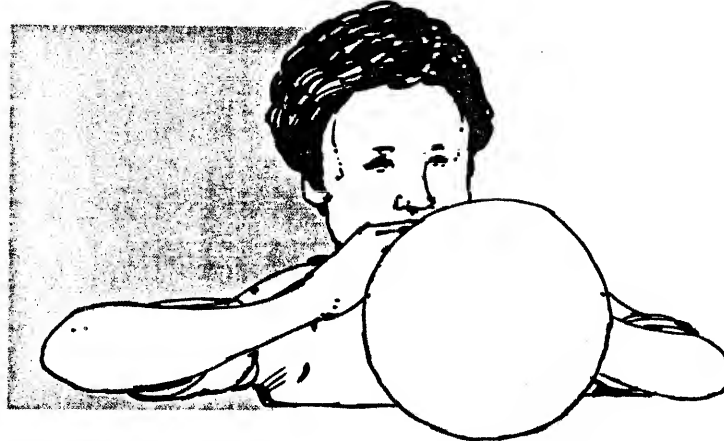
Figure 7-4

- E. Now refill the jug with water and have your partner turn it upside down in the bucket. Insert the tubing as before. This time, take just one really deep breath, as deep as you can, and blow it all out into the tubing. Measure the volume of this deep breath to the nearest 100 ml.



- ✓ 3-3. What was the volume of your deep breath?

Compare your answers to Questions 3-2 and 3-3. Surprised at the difference? Most people can breathe out a volume of air about seven times the volume of a normal breath. Even so, no matter how hard you try to empty your lungs completely, some air will be left. Your total lung capacity is about ten times the volume of a normal breath.



What causes air to move in and out of your lungs? Many people think it's caused by the ribs moving in and out. But this isn't the whole story.

14 CORE